Regularization Approaches for Synthesizing HRTF Directivity Patterns

Eugen Rasumow, Martin Hansen, Steven van de Par, Dirk Püschel, Volker Mellert, Simon Doclo, and Matthias Blau

***Abstract*—As an alternative to traditional artificial heads, it is possible to synthesize individual head-related transfer functions (HRTFs) using a so-called virtual artificial head (VAH), consisting of a microphone array with an appropriate topology and filter coefficients optimized using a narrowband least squares cost function. The resulting spatial directivity pattern of such a VAH is known to be sensitive to small deviations of the assumed microphone characteristics, e.g., gain, phase and/or the positions of the microphones. In many beamformer design procedures, this sensitivity is reduced by imposing a white noise gain (WNG) constraint on the filter coefficients for a single desired look direction. In this paper, this constraint is shown to be inappropriate for regularizing the HRTF synthesis with multiple desired directions and three alternative different regularization approaches are proposed and evaluated. In the first approach, the measured deviations of the microphone characteristics are taken into account in the filter design. In the second approach, the filter coefficients are regularized using the mean WNG for all directions. The third approach additionally takes into account several frequency bins into both the optimization and the regularization. The different proposed regularization approaches are compared using analytic and measured transfer functions, including random deviations. Experimental results show that the approach using multiple frequency bands mimicking the spectral resolution of the human auditory system yields the best robustness among the considered regularization approaches.**

***Index Terms*—Beamforming, head-related transfer functions (HRTFs), regularization, virtual artificial head, white noise gain (WNG).**

I. Introduction

BINAURAL sound reproduction is an important reproduction method aiming to preserve the spatial information, where the goal is to reproduce the sound at the listener's ears via headphones in the same way as if the listener had been in the real sound field. The recordings needed for binaural sound reproduction are traditionally made using so-called artificial heads with average anthropometric characteristics. Unfortunately, because of their non-individual character, these artificial head recordings often entail perceptual deficiencies [1]–[3].

As an alternative to traditional artificial heads, microphone arrays can be used for spatial filtering, e.g., using filter-and-sum beamforming, [4], [5]. Even though many beamformers (e.g., superdirective beamformers [6]–[10]) aim at steering into one look direction, they can, in principle, be used to synthesize an arbitrary desired directivity pattern, e.g., based on different cost functions [11]–[13]. Filter-and-sum beamforming can therefore also be used to synthesize individual head-related transfer functions (HRTFs), thus mimicking the directivity patterns of artificial or real human heads [14]–[23]. This approach is referred to as a virtual artificial head (VAH). The main advantages of a VAH are the possibility to adjust the filter coefficients to HRTFs of different listeners (individualization), the possibility to employ head tracking in the reproduction stage and a better flexibility and manageability due to the smaller size/weight of the device.

However, synthesizing spatial directivity patterns with many microphones and rather small inter-microphone distances is known to be sensitive ([19], [24]) to small deviations of the assumed microphone characteristics (e.g., gain, phase, positions, temperature changes and/or drifting microphone characteristics) [6], [9], [19], [24], [25]. To improve the robustness, some kind of regularization is usually employed, where obviously a trade-off between synthesis accuracy (in case of no deviations) and robustness (in case of deviations) exists. In general, different regularization approaches have been proposed in the literature, cf. [28] for a comprehensive review. A popular regularization approach is to impose a constraint on the filter coefficients, also referred to as Tikhonov regularization [26], whereas other regularization approaches are based on, e.g., the truncated singular value decomposition or the L-curve [27], which provides an optimal trade-off between the cost function and the norm of the filter coefficients.

In this paper, three regularization approaches for synthesizing multi-directional directivity patterns, such as HRTFs, based on a narrowband least squares cost function are proposed and evaluated. In the first approach, the filter coefficients are regularized by taking into account measured deviations of the steering vectors (cf. Section III), which is comparable to taking into account the probability density function of the microphone characteristics [9], [24]. Since usually such measured deviations of the steering vectors are not available, it is common practice to impose a so-called white noise gain (WNG) constraint on the filter coefficients [6], [8], [10], however typically only for the desired

Manuscript received June 15, 2015; revised October 19, 2015; accepted November 15, 2015. Date of publication December 03, 2015; date of current version January 04, 2016. This work was supported in part by Bundesministerium für Bildung und Forschung under Grant 17080X10, in part by the Cluster of Excellence 1077 "Hearing4All" of the German Research Foundation, and in part by Akustik Technologie Göttingen. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Woon-Seng Gan.

E. Rasumow, M. Hansen, and M. Blau are with the Institute of Hearing Technology and Audiology, Jade Hochschule Oldenburg, D-26121 Oldenburg, Germany (e-mail: eugen.rasumow@web.de).

S. van de Par, V. Mellert, and S. Doclo are with the Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, University of Oldenburg, D-26111 Oldenburg, Germany.

D. Püschel is with the Akustik Technologie Göttingen, D-37037 Göttingen, Germany.

Digital Object Identifier 10.1109/TASLP.2015.2504874

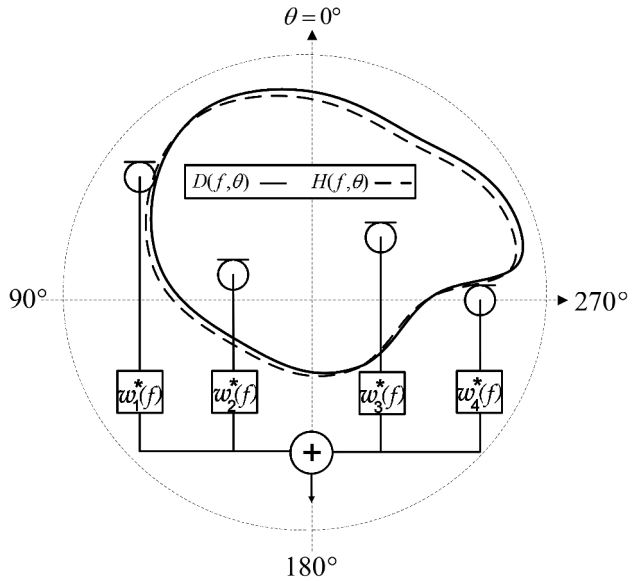


Fig. 1. Schematic diagram of a filter-and-sum beamformer with $N = 4$ microphones and an exemplary desired directivity pattern $\mathrm{D}(f,\theta)$ and resulting directivity pattern $\mathrm{H}(f,\theta)$.

look direction. In this paper, it is shown in Section V-C that this constraint is inappropriate for synthesizing HRTFs with multiple desired directions. Hence, in the second approach a different WNG constraint is proposed, incorporating all directions into the weighting of the WNG (cf. Section IV-B). In the third approach, the information of neighboring frequency bands is additionally considered in the optimization and regularization procedure (cf. Section IV-C). In Section V, these different approaches to increase the robustness of the VAH synthesis are evaluated with measured and analytic steering vectors, with and without adding random deviations.

The perceptual suitability of the different regularization approaches to synthesize HRTFs using a VAH has already been separately shown in [22], [23], [37] and [38]. The main objective of this paper is to evaluate the robustness of these approaches against deviations of the steering vectors using physical performance measures.

II. Least Squares Beamformer Design

In general, to synthesize a desired directivity pattern using a microphone array, the filter coefficients of a filter-and-sum beamformer (cf. Fig. 1) can be computed by minimizing a cost function (e.g., least squares, total least squares or non-linear cost functions [12], [13]), either for all frequencies jointly (broadband design) or for each frequency independently (narrowband design). In this paper, a narrowband design procedure is used because of its better numerical stability and since previous studies have already shown its suitability to synthesize HRTFs using a VAH [22], [23].

The synthesized spatial directivity pattern $\mathrm{H}(f,\theta)$ a filter-and-sum beamformer can be expressed as[1]

$$\mathrm{H}(f,\theta) = \mathbf{w}^H(f)\mathbf{d}(f,\theta), \tag{1}$$

[1]In the following $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$, $\mathbf{x}^H$ denotes the Hermitian transpose of $\mathbf{x}$ and x^* denotes the complex conjugate of x.

with $\mathbf{d}(f,\theta)$ the N-dimensional steering vector, describing the acoustic transfer functions between a sound source from direction[2] θ to the N microphones at frequency f, and $\mathbf{w}(f)$ an N-dimensional vector containing the complex-valued filter coefficients, i.e. $\mathbf{w}(f) = [w_1(f), w_2(f) \ldots w_N(f)]^T$. In order to synthesize a desired directivity pattern $\mathrm{D}(f,\theta)$, e.g., an individual frequency- and direction-dependent HRTF (cf. Fig. 1), the filter coefficients $\mathbf{w}(f)$ can be computed by minimizing the narrowband weighted least squares cost function,

$$\mathrm{J_{LS}}(\mathbf{w}(f)) = \sum_{i=1}^{\mathrm{P}} F(f,\theta_i) \cdot |\underbrace{\mathbf{w}^H(f)\mathbf{d}(f,\theta_i)}_{\mathrm{H}(f,\theta_i)} - \mathrm{D}(f,\theta_i)|^2 \tag{2}$$

i.e. the weighted sum over P discrete directions of the squared absolute difference between the synthesized directivity pattern $\mathrm{H}(f,\theta_i)$ and the desired directivity pattern $\mathrm{D}(f,\theta_i)$. The weighting function $F(f,\theta)$ enables to assign more or less importance to certain directions. In the remainder of this paper, the least squares cost function in (2) will be used because it has a closed-form solution and since previous studies have already shown its suitability to synthesize HRTFs [22], [23].

The filter coefficients minimizing this cost function can be obtained by setting the gradient $\nabla_{\mathbf{w}(f)}\mathrm{J_{LS}}(\mathbf{w}(f))$ to zero, leading to

$$\mathbf{w}_{\mathrm{LS}}(f) = \mathbf{Q}^{-1}(f) \cdot \mathbf{a}(f), \tag{3}$$

with

$$\mathbf{Q}(f) = \sum_{i=1}^{\mathrm{P}} F(f,\theta_i) \cdot \mathbf{d}(f,\theta_i)\mathbf{d}^H(f,\theta_i),$$

$$\mathbf{a}(f) = \sum_{i=1}^{\mathrm{P}} F(f,\theta_i) \cdot \mathbf{d}(f,\theta_i)\mathrm{D}^*(f,\theta_i).$$

When using many microphones with small inter-microphone distances, the synthesized directivity pattern $\mathrm{H}(f,\theta)$ is known to be highly sensitive to small deviations of the assumed steering vectors $\mathbf{d}(f,\theta)$. In the next sections, different regularization approaches will be discussed, either by using multiple measured steering vectors or by imposing an appropriate constraint on the filter coefficients $\mathbf{w}(f)$.

III. Regularization by Joint Optimization for Multiple Steering Vectors

When multiple sets of steering vectors are available, e.g., measured under different conditions (e.g., temperature, different microphone positions), the beamformer robustness can be increased by jointly optimizing the least squares cost function in (2) over all available sets of steering vectors, i.e.

$$\mathrm{J_K}(\mathbf{w}(f)) = \sum_{k=1}^{\mathrm{K}} \sum_{i=1}^{\mathrm{P}} F_k(f,\theta_i)\left|\mathbf{w}^H(f)\mathbf{d}_k(f,\theta_i) - \mathrm{D}(f,\theta_i)\right|^2 \tag{4}$$

[2]In this paper only the azimuthal direction θ is considered, but the proposed procedure can be straightforwardly extended to a three-dimensional design including elevation directions.

with K the total number of measured sets of steering vectors $\mathbf{d}_k$ and $F_k(f,\theta)$ the weighting function for the k-th set of steering vectors. Analogous to (3), the filter coefficients minimizing (4) are equal to

$$\mathbf{w}_{\mathrm{K}}(f) = \mathbf{Q}_{\mathrm{K}}^{-1}(f) \cdot \mathbf{a}_{\mathrm{K}}(f), \tag{5}$$

with

$$\mathbf{Q}_{\mathrm{K}}(f) = \sum_{k=1}^{\mathrm{K}} \sum_{i=1}^{\mathrm{P}} F_k(f,\theta_i) \cdot \mathbf{d}_k(f,\theta_i)\mathbf{d}_k^H(f,\theta_i)$$
$$\mathbf{a}_{\mathrm{K}}(f) = \sum_{k=1}^{\mathrm{K}} \sum_{i=1}^{\mathrm{P}} F_k(f,\theta_i) \cdot \mathbf{d}_k(f,\theta_i)\mathrm{D}^*(f,\theta_i).$$

This joint optimization can be interpreted as a regularization using measured data and is comparable to incorporating assumed deviations of the microphone characteristics as proposed in [9], [24]. In contrast to using probability density functions as in [9], [24], the measured deviations of the steering vectors may be considered as an empirical estimate for the expected deviations of the steering vectors.

However, it should be realized that in order to capture all possible real-world deviations of the steering vectors, a large number of sets of steering vectors would need to be measured, making this approach rather impractical. In addition, when only a limited set of measured steering vectors is available, the robustness of the beamformer can only be improved to a limited extent (cf. Section V-F).

IV. Regularization using a WNG Constraint

Since typically only one set of measured steering vectors is available, it is common practice to improve the robustness of the beamformer by imposing a constraint on the filter coefficients. Due to its physical meaning, the output power of the beamformer for the desired acoustic field in comparison to the output power for spatially uncorrelated noise, defined as the white noise gain (WNG), is a common measure to quantify the robustness of a beamformer [6], [8]. The main motivation of regularization by constraining the WNG is to reduce the output power for spatially uncorrelated noise (in comparison to the output power for the desired acoustic field), aiming to enhance the overall robustness. However, as will be shown, the definition of the desired acoustic field has a significant influence on the resulting performance. In the following sections, the advantages and disadvantages of various WNG variants with regard to their application for a VAH will be discussed.

A. White Noise Gain

The frequency- and direction-dependent white noise gain is defined as [8]

$$\mathrm{WNG}(\mathbf{w}(f),\theta) = \frac{\left|\mathbf{w}^H(f)\mathbf{d}(f,\theta)\right|^2}{\mathbf{w}^H(f)\mathbf{w}(f)}. \tag{6}$$

For many beamformers, e.g., superdirective beamformers [6]–[8], the WNG is considered only for the desired look direction θ_d. In this case, $\mathrm{WNG}(\mathbf{w}(f),\theta_d)$ relates the output power for the look direction θ_d to the output power for spatially uncorrelated noise. In general, a larger WNG is associated with a larger attenuation of spatially uncorrelated noise in comparison to the output power for the look direction θ_d and hence with an increased robustness, cf. [6]–[9].

When imposing a constraint on the WNG for the direction θ_d, the constrained optimization problem can be written as

$$\boxed{\min_{\mathbf{w}(f)} \mathrm{J}_{\mathrm{LS}}(\mathbf{w}(f)) \quad \text{subject to} \quad \mathrm{WNG}(\mathbf{w}(f),\theta_d) \geq \beta} \tag{7}$$

with β the minimum desired value for $\mathrm{WNG}(\mathbf{w}(f),\theta_d)$. Using (6), the Lagrangian function associated with this constrained optimization problem is equal to

$$\mathrm{J}_{\mathrm{sd}}(\mathbf{w}(f),\mu) = \mathrm{J}_{\mathrm{LS}}(\mathbf{w}(f)) + \mu\left(\mathbf{w}^H(f)\mathbf{w}(f) - \frac{1}{\beta}\left|\mathbf{w}^H(f)\mathbf{d}(f,\theta_d)\right|^2\right), \tag{8}$$

with μ the Lagrangian multiplier. Analogous to (3) and (5), the filter coefficients minimizing the cost function in (8) can be obtained by setting the gradient to zero, leading to

$$\mathbf{w}_{\mathrm{sd}}(f,\mu) = \left(\mathbf{Q}(f) + \mu\left(\mathbf{I}_N - \frac{1}{\beta}\mathbf{d}(f,\theta_d)\mathbf{d}^H(f,\theta_d)\right)\right)^{-1}\mathbf{a}(f), \tag{9}$$

with $\mathbf{Q}(f)$ and $\mathbf{a}(f)$ given in (3) and $\mathbf{I}_N$ the $N \times N$-dimensional identity matrix. In order to satisfy the inequality constraint $\mathrm{WNG}(\mathbf{w}_{\mathrm{sd}}(f,\mu),\theta_d) \geq \beta$, the Lagrange multiplier μ in (9) needs to be determined, e.g., using an iterative procedure (cf. Section V-A for a detailed description of the used procedure).

Note that if the beamformer response in the look direction $\mathrm{H}(f,\theta_d) = \mathbf{w}^H(f)\mathbf{d}(f,\theta_d)$ is equal to 1, which is typically the case for superdirective beamformers [6]–[8], the WNG in (6) reduces to

$$\mathrm{WNG}(\mathbf{w}(f),\theta_d) = \frac{1}{\mathbf{w}^H(f)\mathbf{w}(f)},$$

such that the term $\frac{1}{\beta}|\mathbf{w}^H(f)\mathbf{d}(f,\theta_d)|^2$ vanishes in (8) and the filter coefficients in (9) are equal to

$$\mathbf{w}_{\mathrm{dl}}(f,\mu) = (\mathbf{Q}(f) + \mu \cdot \mathbf{I}_N)^{-1}\mathbf{a}(f). \tag{10}$$

The experimental results in Section V will show that imposing a WNG constraint for a single look direction θ_d is inappropriate for synthesizing HRTFs, where the objective is to synthesize a desired directivity pattern for multiple directions. This has also been confirmed by the perceptual validations in [22]. Hence, a direction-dependent weighting for $\mathrm{WNG}(\mathbf{w}(f),\theta)$ will be presented in the following sections.

B. Mean White Noise Gain Over All Directions

When synthesizing multi-directional directivity patterns such as HRTFs, the accuracy and the robustness of the synthesized directivity pattern needs to be assured for all considered directions. Hence, instead of constraining the WNG in (6) for only

one direction, we propose to constrain the mean white noise gain over all directions, defined as

$$\boxed{\mathrm{WNG_m}(\mathbf{w}(f)) = \sum_{i=1}^{\mathrm{P}} g(f,\theta_i)\ \mathrm{WNG}(\mathbf{w}(f),\theta_i)} \tag{11}$$

where the weighting function $g(f,\theta)$ enables to assign more or less importance to certain directions [23]. Using (6), the mean white noise gain in (11) can be rewritten as

$$\mathrm{WNG_m}(\mathbf{w}(f)) = \frac{\mathbf{w}^H(f)\mathbf{Q}_\mathrm{m}(f)\ \mathbf{w}(f)}{\mathbf{w}^H(f)\mathbf{w}(f)}, \tag{12}$$

with

$$\mathbf{Q}_\mathrm{m}(f) = \sum_{i=1}^{\mathrm{P}} g(f,\theta_i)\cdot \mathbf{d}(f,\theta_i)\mathbf{d}^H(f,\theta_i). \tag{13}$$

In the remainder of this paper, we will use a uniform weighting for all directions, i.e. $g(f,\theta_i) = \frac{1}{\mathrm{P}}, i = 1\ldots\mathrm{P}$. In this case, $\mathrm{WNG_m}(\mathbf{w}(f))$ relates the mean output power from all P considered directions to the output power for spatially uncorrelated noise.

When imposing a constraint on the mean white noise gain, the constrained optimization problem can be written as

$$\boxed{\min_{\mathbf{w}(f)}\ \mathrm{J_{LS}}(\mathbf{w}(f)) \quad \text{subject to} \quad \mathrm{WNG_m}(\mathbf{w}(f)) \geq \beta_\mathrm{m}} \tag{14}$$

with β_m the minimum desired value for $\mathrm{WNG_m}$. The Lagrangian function associated with this constrained optimization problem is equal to

$$\mathrm{J_m}(\mathbf{w}(f),\mu) = \mathrm{J_{LS}}(\mathbf{w}(f)) + \mu\left(\mathbf{w}^H(f)\mathbf{w}(f) - \frac{1}{\beta_\mathrm{m}}\ \mathbf{w}^H(f)\mathbf{Q}_\mathrm{m}(f)\mathbf{w}(f)\right). \tag{15}$$

Analogous to (9), the filter coefficients minimizing the cost function in (15) are equal to

$$\mathbf{w}_\mathrm{m}(f,\mu) = \left(\mathbf{Q}(f) + \mu\left(\mathbf{I}_N - \frac{1}{\beta_\mathrm{m}}\mathbf{Q}_\mathrm{m}(f)\right)\right)^{-1}\mathbf{a}(f). \tag{16}$$

Hence, the only difference between the solutions in (9) and (16) is the exchange of the rank-1 matrix $\mathbf{d}(f,\theta_d)\mathbf{d}^H(f,\theta_d)$ in (9) by the rank-P (assuming independent steering vectors from P directions) matrix $\mathbf{Q}_\mathrm{m}(f)$ in (16).

Note that this regularization approach has been perceptually validated in [23] in terms of localization performance, sensor noise, spectral coloration and overall performance, where it has been shown that the overall performance for all evaluated directions range from fair to excellent for all subjects.

C. Optimization and White Noise Gain for Multiple Frequencies

Since it is a well-known phenomenon that the human auditory system groups incoming sounds into so-called critical frequency bands that broaden with increasing center frequencies [29], in this section we propose a cost function and a WNG constraint that incorporate the grouping of frequencies within a perceptually-relevant bandwidth. Consequently, the filter optimization at each frequency f is formulated within equivalent rectangular bandwidths (ERB), corresponding to human auditory filters [30], with f as its center frequency.

Let us consider the t-th frequency band with center frequency f_c^t, containing L^t frequency bins. We now define the frequency vector $\mathbf{\Omega}^t = f_1^t \ldots f_c^t \ldots f_{L^t}^t$, where f_1^t and $f_{L^t}^t$ denote the first and the last frequency bin. Furthermore, we define the $N\cdot L^t$-dimensional stacked filter vector $\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t)$ for the t-th frequency band as

$$\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t) = \begin{bmatrix} \mathbf{w}(f_1^t) \\ \vdots \\ \mathbf{w}(f_c^t) \\ \vdots \\ \mathbf{w}(f_{L^t}^t) \end{bmatrix}. \tag{17}$$

The cost function for the t-th frequency band can be defined as the sum of the least squares cost functions in the L^t frequency bins, i.e.

$$\begin{aligned}\mathrm{J_v}(\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t)) &= \sum_{l=1}^{L^t} \mathrm{J_{LS}}(\mathbf{w}(f_l^t)) \\ &= \sum_{l=1}^{L^t}\sum_{i=1}^{\mathrm{P}} F(f_l^t,\theta_i)\cdot\left|\mathbf{w}^H(f_l^t)\mathbf{d}(f_l^t,\theta_i) - \mathrm{D}(f_l^t,\theta_i)\right|^2,\end{aligned} \tag{18}$$

which can be rewritten using the stacked filter vector in (17) as

$$\begin{aligned}\mathrm{J_v}(\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t)) = \mathbf{w}_\mathrm{v}^H(\mathbf{\Omega}^t)\mathbf{Q}_\mathrm{v}(\mathbf{\Omega}^t)\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t) - \mathbf{w}_\mathrm{v}^H(\mathbf{\Omega}^t)\mathbf{a}_\mathrm{v}(\mathbf{\Omega}^t) \\ - \mathbf{a}_\mathrm{v}^H(\mathbf{\Omega}^t)\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t) + d_\mathrm{v}(\mathbf{\Omega}^t),\end{aligned} \tag{19}$$

with

$$\mathbf{Q}_\mathrm{v}(\mathbf{\Omega}^t) = \begin{bmatrix} \mathbf{Q}(f_1^t) & & & & \\ & \ddots & & & \\ & & \mathbf{Q}(f_c^t) & & \\ & & & \ddots & \\ & & & & \mathbf{Q}(f_{L^t}^t) \end{bmatrix},$$

$$\mathbf{a}_\mathrm{v}(\mathbf{\Omega}^t) = \begin{bmatrix} \mathbf{a}(f_1^t) \\ \vdots \\ \mathbf{a}(f_c^t) \\ \vdots \\ \mathbf{a}(f_{L^t}^t) \end{bmatrix},$$

$$d_\mathrm{v}(\mathbf{\Omega}^t) = \sum_{l=1}^{L^t}\sum_{i=1}^{\mathrm{P}} F(f_l^t,\theta_i)\cdot\left|\mathrm{D}(f_l^t,\theta_i)\right|^2. \tag{20}$$

Similarly to (12), the WNG for the t-th frequency band can be defined as

$$\mathrm{WNG_v}(\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t)) = \frac{\sum_{l=1}^{L^t}\mathbf{w}^H(f_l^t)\mathbf{Q}_\mathrm{m}(f_l^t)\mathbf{w}(f_l^t)}{\sum_{l=1}^{L^t}\mathbf{w}^H(f_l^t)\mathbf{w}(f_l^t)}, \tag{21}$$

which relates the mean output power from all P directions summed over the L^t considered frequency bins in $\mathbf{\Omega}^t$ to the

output power for spatially uncorrelated noise summed over the L^t considered frequency bins. The WNG in (21) can be rewritten using the stacked filter vector in (17) as

$$\mathrm{WNG_v}(\mathbf{w_v}(\mathbf{\Omega}^t)) = \frac{\mathbf{w}_\mathrm{v}^H(\mathbf{\Omega}^t)\mathbf{Q}_\mathrm{vm}(\mathbf{\Omega}^t)\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t)}{\mathbf{w}_\mathrm{v}^H(\mathbf{\Omega}^t)\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t)}, \qquad (22)$$

with

$$\mathbf{Q}_\mathrm{vm}(\mathbf{\Omega}^t) = \begin{bmatrix} \mathbf{Q}_\mathrm{m}(f_1^t) & & & \\ & \ddots & & \\ & & \mathbf{Q}_\mathrm{m}(f_c^t) & \\ & & & \ddots \\ & & & \mathbf{Q}_\mathrm{m}(f_{L^t}^t) \end{bmatrix}. \qquad (23)$$

When imposing a constraint on the WNG for the t-th frequency band, the constrained optimization problem for the t-th frequency band can be written as

$$\boxed{\min_{\mathbf{w_v}(\mathbf{\Omega}^t)} \mathrm{J_v}(\mathbf{w_v}(\mathbf{\Omega}^t)) \quad \text{subject to} \quad \mathrm{WNG_v}(\mathbf{w_v}(\mathbf{\Omega}^t)) \geq \beta_\mathrm{v}} \qquad (24)$$

with β_v the minimum desired value for $\mathrm{WNG_v}$. The Lagrangian function associated with this constrained optimization problem is equal to

$$\mathrm{J_{vm}}(\mathbf{w_v}(\mathbf{\Omega}^t), \mu) = \mathrm{J_v}(\mathbf{w_v}(\mathbf{\Omega}^t)) + \mu\Big(\mathbf{w}_\mathrm{v}^H(\mathbf{\Omega}^t)\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t) - \frac{1}{\beta_\mathrm{v}}\mathbf{w}_\mathrm{v}^H(\mathbf{\Omega}^t)\mathbf{Q}_\mathrm{vm}(\mathbf{\Omega}^t)\mathbf{w}_\mathrm{v}(\mathbf{\Omega}^t)\Big). \qquad (25)$$

Analogous to (16), the filter coefficients minimizing $\mathrm{J_{vm}}$ are equal to

$$\mathbf{w}_\mathrm{vm}(\mathbf{\Omega}^t, \mu) = \left(\mathbf{Q}_\mathrm{v}(\mathbf{\Omega}^t) + \mu\left(\mathbf{I}_\mathrm{v} - \frac{1}{\beta_\mathrm{v}}\mathbf{Q}_\mathrm{vm}(\mathbf{\Omega}^t)\right)\right)^{-1}\mathbf{a}_\mathrm{v}(\mathbf{\Omega}^t), \qquad (26)$$

with $\mathbf{I}_\mathrm{v}$ the $N \cdot L^t \times N \cdot L^t$-dimensional identity matrix.

It is important to note that we solve the optimization problem in (24) for all frequency bins (and not only for the center frequencies of the critical auditory bands), i.e. for each frequency f we consider the ERB with frequencies $\mathbf{\Omega}^t$ around that frequency as its center frequency f_c^t. The solution in (26) then yields an $N \cdot L^t$-dimensional vector with filter coefficients $\mathbf{w}(f_1^t) \ldots \mathbf{w}(f_{L^t}^t)$, where we only consider the filter coefficients $\mathbf{w}(f_c^t)$ at the center frequency as the solution for that frequency bin. This procedure can hence be interpreted as taking neighboring frequencies into account both for the cost function in (18) as well as for the regularization in (21). It is worth noting that for $\mu = 0$, i.e. without regularization, the solution in (16), taking into account a single frequency band, and the solution in (26), taking into account neighboring frequency bands, yield very similar numerical results.

In contrast to the WNG for a single frequency bin in (12), for the WNG in (21) the weighting function $g(f, \theta)$ in (13) can be chosen to assign more or less importance to certain directions and frequencies. In the remainder of this paper, we will use a uniform weighting for all directions and frequencies, i.e. $g(f_l^t, \theta_i) = \frac{1}{\mathrm{P}}, i = 1 \ldots \mathrm{P}, l = 1 \ldots L^t$.

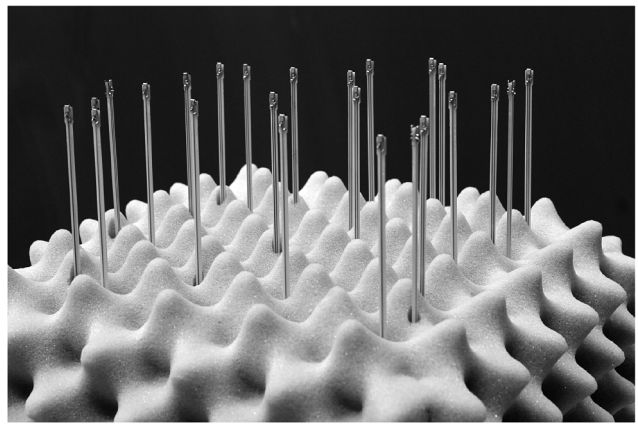


Fig. 2. Used planar microphone array with N = 24 microphones with a topology according to [19].

Note that this regularization approach has been perceptually validated in [37], [38] in terms of localization, spectral coloration and overall performance, where it has been shown that the median perceptual ratings for the VAH range from good to excellent and are in general better than the median perceptual ratings for traditional artificial heads.

V. Experimental Results

In this section, the proposed optimization and regularization approaches from Sections III and IV are compared with respect to their synthesis accuracy in case of no deviations (Sections V-B–V-E) and their robustness against random deviations of the steering vectors (Section V-F).

In order to quantify the *accuracy* of the synthesized directivity pattern, we use the logarithmic error

$$\epsilon(f_c, \theta) = \frac{1}{L}\sum_{l=1}^{L}\left|20\log_{10}\left(\left|\frac{\mathrm{D}(f_l, \theta)}{\mathrm{H}(f_l, \theta)}\right|\right)\right|, \qquad (27)$$

where for each direction θ the mean absolute dB-error is computed for all frequencies in the ERB-band centered around its center frequency f_c with a 50% overlap [31]. Note that (27) is used as the accuracy measure instead of the least squares error in (2) because it is better suited to represent the perceptually relevant error of the synthesis [31]. This measure was, however, not used as a cost function because no closed-form solution for minimizing (27) exists.

In order to quantify the *robustness* of the different approaches, we use both the mean white noise gain $\mathrm{WNG_m}$ (cf. Section V-B) and the mean synthesis error for randomly disturbed steering vectors (cf. Section V-F).

A. Measurement Setup and Algorithmic Parameters

The HRTFs, i.e. the desired directivity pattern $\mathrm{D}(f, \theta)$, were measured in the horizontal plane using the blocked ear method [33], where microphones (Knowles FG-23329 miniature electret microphones) were flush mounted in individualized earmolds that blocked the ear entrance. In this paper we will use the left ear HRTFs of subject $\mathrm{S_1}$ (for details we refer to [31]). The steering vectors $\mathbf{d}(f, \theta)$ were measured with a planar microphone array (cf. Fig. 2) with a Golomb-based topology, which is a mathematically-motivated method for

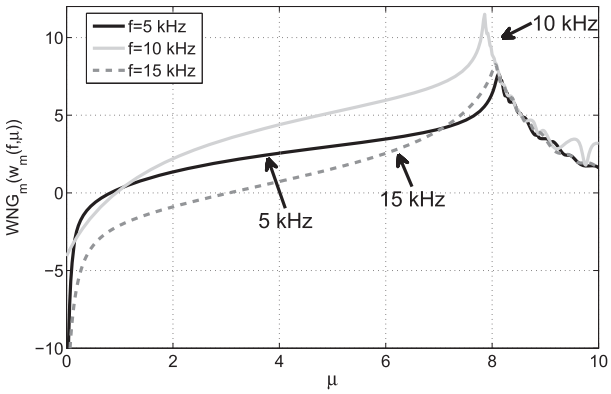


Fig. 3. Mean white noise gain $\mathrm{WNG_m}$ as a function of the Lagrange multiplier μ for three frequencies.

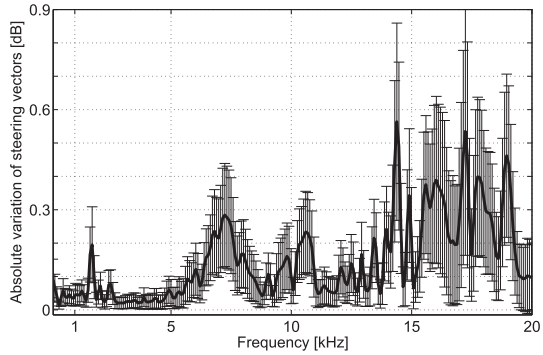


Fig. 4. Mean (black line) and standard deviation (error bar) of the variability (absolute values) among the different measured steering vectors (note the linear frequency scaling).

deriving microphone array topologies [19]. The microphone array consisted of $N = 24$ microphones (each consisting of two Analog Devices ADMP 504 Ultralow Noise sensors) on a 20 cm $\times$ 20 cm plate covered with additional absorbing material. The steering vectors and the HRTFs were measured in an anechoic room in the horizontal plane with an azimuthal resolution of $\Delta\theta = 15°$, i.e. $\mathrm{P} = 24$ directions[3]. This resolution was shown to be sufficient to synthesize accurate sound space information [37] and hence was chosen for the objective comparison of the regularization approaches. Both the steering vectors and the HRTFs were measured with white noise stimuli using the $\mathrm{H_1}$ estimate [32] with an 8192-point Hann window, 50% overlap and 27 averages, and were truncated in the time domain to a length of 256 samples (corresponding to about 5.8 ms at a sampling frequency of 44.1 kHz). The HRTFs were smoothed (according to the perceptual limits derived in [31]) into constant relative bandwidths of $\mathrm{B_W} = \frac{1}{5}$ octaves in the frequency domain and the spatial notches of the HRTF directivity patterns were levelled out per frequency such that the dynamic range of the directivity patterns across azimuth did not exceed 29 dB. Furthermore, the measured steering vectors were normalized by a constant factor

$$\nu_k = \sqrt{\frac{1}{N \cdot L \cdot \mathrm{P}} \sum_{l=1}^{L} \sum_{i=1}^{\mathrm{P}} \mathbf{d}_k^H(f_l, \theta_i)\mathbf{d}_k(f_l, \theta_i)}, \qquad (28)$$

with $f_1 = 500$ Hz, $f_L = 1000$ Hz and L the number of intermediate frequency bins, in order to achieve unit power on average over all directions, microphones and frequencies 500 Hz $\leq f \leq$ 1000 Hz.

For each of the discussed regularization approaches, the Lagrange multiplier μ in (9), (16) and (26) was determined numerically. To determine the smallest possible μ (resulting in the most accurate synthesis) satisfying the desired inequality constraint in (7), (14) and (24), μ was increased logarithmically until the resulting WNG reached the desired WNG within an accuracy of 0.05 dB. For example, Fig. 3 depicts the mean white noise gain $\mathrm{WNG_m}$ for three frequencies, where it can be observed that the mean white noise gain does not always increase monotonically with increasing μ. The minimum desired WNG values in (7), (14) and (24) were set to $\beta = \beta_{\mathrm{m}} = \beta_{\mathrm{v}} = -5$ dB. This desired WNG value was chosen based on the perceptual study in [23], where it was shown that for the used microphone array the best performance in terms of localization performance, sensor noise, spectral coloration and overall performance was obtained for -6 dB $\leq \beta_{\mathrm{m}} \leq 0$ dB.

B. Joint Optimization for Multiple Steering Vectors

To demonstrate the effect of a joint optimization for multiple sets of steering vectors (cf. Section III), $\mathrm{K} = 4$ sets of steering vectors were measured with the same measuring apparatus (with approximately 24 hours lying between the measurements). These measured sets of steering vectors were integrated into a joint optimization according to (4) with a uniform weighting, i.e. $F_k(f, \theta) = 1$.

In Fig. 4, the mean and the standard deviation of the variability among the sets of measured steering vectors (absolute values) is illustrated. Although the positioning and the environmental influences of the measuring apparatus were kept as constant as possible, the absolute value of the measured transfer functions clearly varied between the measurements. Explanations for these variations may be, e.g., drifting characteristics of the measuring apparatus and/or slightly differing environmental conditions, such as temperature [34], [35]. These measured deviations may be considered as an empirical estimate of the expected deviations of the steering vectors, although clearly not all possible real-world deviations of the steering vectors will be captured with this limited amount of measurement sets.

As a measure for robustness, the mean white noise gain $\mathrm{WNG_m}$ for the (unconstrained) joint optimization using $\mathrm{K} = 1 \ldots 4$ sets of steering vectors is shown in Fig. 5. It can be observed that the filter coefficients resulting from an unconstrained optimization with $\mathrm{K} = 1$ set of steering vectors yield very low $\mathrm{WNG_m}$ (corresponding to a low robustness), down to $\mathrm{WNG_m} \approx -85$ dB (not visible in Fig. 5) at lower frequencies ($f \leq 4$ kHz). It can also be clearly observed that the $\mathrm{WNG_m}$ increases with increasing K, demonstrating that it is possible to increase the robustness by using multiple sets of

[3]It should be noted that in the presented experiments the number of directions P and the number of microphones N was equal (i.e. $\mathrm{P} = \mathrm{N}$). However, based on informal listening tests the performance of the synthesis does not decrease drastically when slightly reducing the number of microphones (i.e. $\mathrm{P} > \mathrm{N}$), which is presumably a consequence of the applied regularization.

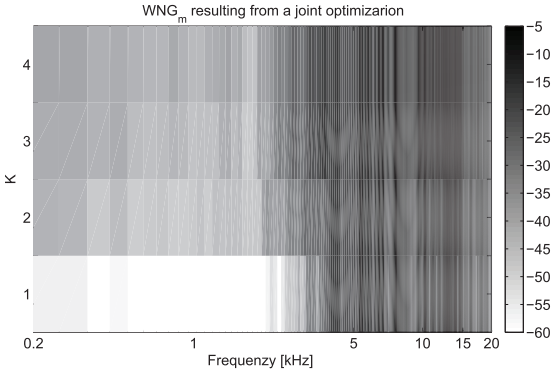


Fig. 5. Mean white noise gain of filter coefficients $\mathbf{w}_{\mathrm{K}}(f)$ resulting from the (unconstrained) joint optimization for different numbers of considered sets of steering vectors K. For the sake of clarity, the range of the depicted $\mathrm{WNG_m}$ is limited between -60 dB $\leq \mathrm{WNG_m} \leq -5$ dB.

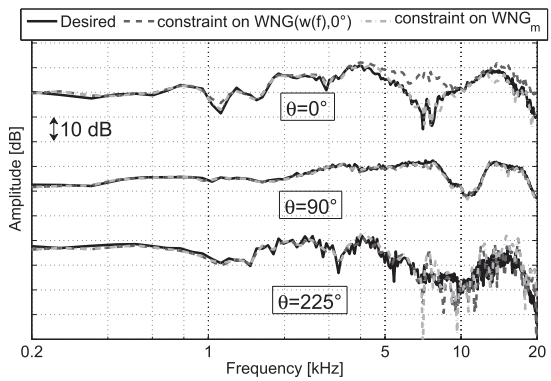


Fig. 6. Desired HRTFs (black solid lines) and synthesized transfer functions associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$ (dot-dashed lines, $\beta = -5$ dB) and on $\mathrm{WNG_m}$ (dashed gray lines, $\beta_{\mathrm{m}} = -5$ dB) for three directions ($\theta = 0°, 90°, 225°$).

measured steering vectors. However, in order to achieve sufficient robustness[4] against different kinds of deviations, a large number of measurement sets is required, which is typically not feasible in practice.

C. White Noise Gain Constraint

Fig. 6 depicts the desired HRTFs and the synthesized transfer functions associated with a constraint on the white noise gain for a single direction ($\theta_d = 0°$) for three directions ($\theta = 0°, 90°, 225°$). The synthesized transfer functions associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$ result in a large deviation from the desired HRTFs for $\theta = \theta_d = 0°$ (especially for 5 kHz $\leq f \leq$ 10 kHz and $f \geq$ 13 kHz). Moreover, the synthesized transfer function associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$ also results in large deviations for $\theta = 225°$ (especially for $f \geq$ 7 kHz), while the synthesized transfer function for $\theta = 90°$ approximates the desired HRTF quite well. The synthesis error ϵ associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$ is depicted in Fig. 8 as a function of frequency and direction. It can be observed that the largest synthesis errors ϵ primarily occur for the frontal direction $\theta = \theta_d = 0°$ and

[4]Meaningful frequency ranges with $\mathrm{WNG_m} \geq 5$ dB are obtained roughly for 12.7 − 13.1 kHz & 13.7 − 14 kHz with K = 1, 12.7 − 13.1 kHz & 13.7 − 14.1 kHz with K = 2, 12.3 − 13.2 kHz & 13.7 − 14.6 kHz with K = 3 and 12.1 − 13.4 kHz & 13.7 − 14.7 kHz with K = 4.

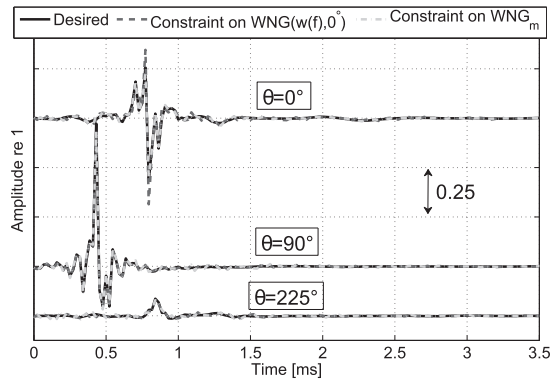


Fig. 7. Desired head-related impulse responses (black solid lines) and synthesized impulse responses associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$ (dot-dashed lines, $\beta = -5$ dB) and on $\mathrm{WNG_m}$ (dashed gray lines, $\beta_{\mathrm{m}} = -5$ dB) for three directions ($\theta = 0°, 90°, 225°$). The impulse responses are separated in amplitude by 0.75 ($\theta = 0°$ and $\theta = 90°$) and by 0.25 ($\theta = 90°$ and $\theta = 225°$).

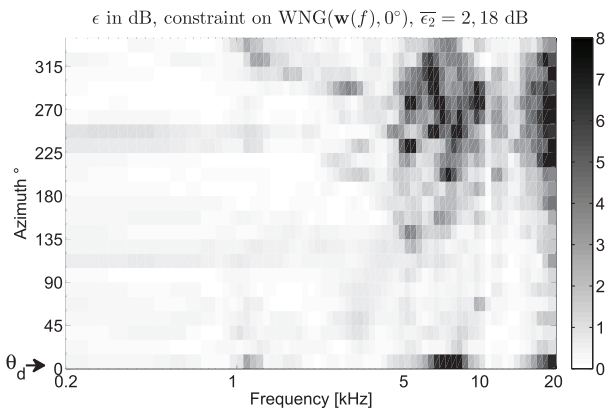


Fig. 8. Synthesis error ϵ as a function of frequency and direction for the synthesis of a left-ear HRTF associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$ ($\beta = -5$ dB). The illustration of ϵ is limited to 8 dB for the sake of clarity.

for contralateral directions ($180° \leq \theta \leq 360°$) and at higher frequencies. Interestingly, the largest errors exactly occur for the direction θ_d which is used for the WNG constraint. This has also been confirmed when changing θ_d to other directions than 0°. This may be explained by the fact that the Lagrange multiplier μ determines the trade-off between synthesis accuracy and robustness. Hence, if a WNG constraint is imposed for a certain direction θ_d, the robustness of the synthesis is enhanced at the expense of decreasing the synthesis accuracy for this direction. Thus, a WNG constraint for a single direction implements a direction-dependent impact on the synthesis, which is clearly an undesirable effect, which has also been validated by the perceptual evaluations in [22].

D. Mean White Noise Gain Constraint

Fig. 6 also depicts the synthesized transfer functions associated with a constraint on the mean white noise gain $\mathrm{WNG_m}$ for three directions ($\theta = 0°, 90°, 225°$). For $\theta = 0$, it can be clearly observed that the synthesized transfer function yields a smaller deviation from the desired HRTF than the synthesized transfer function associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$. On the other hand, for the contralateral direction $\theta = 225$ the synthesized transfer function exhibits larger deviations from the desired HRTF at frequencies $f \geq 4$ kHz.

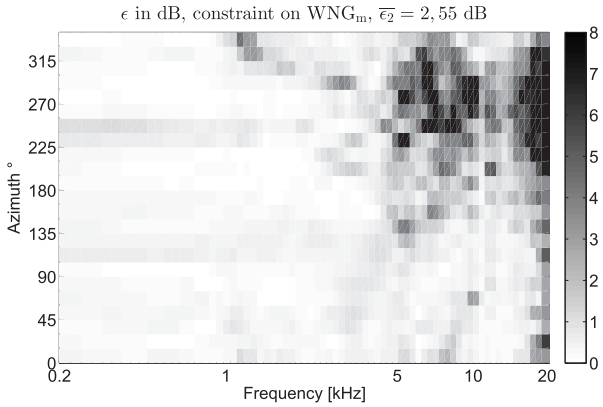


Fig. 9. Synthesis error ϵ as a function of frequency and direction for the synthesis of a left-ear HRTF associated with a constraint on $\mathrm{WNG_m}$ ($\beta_\mathrm{m} = -5$ dB).

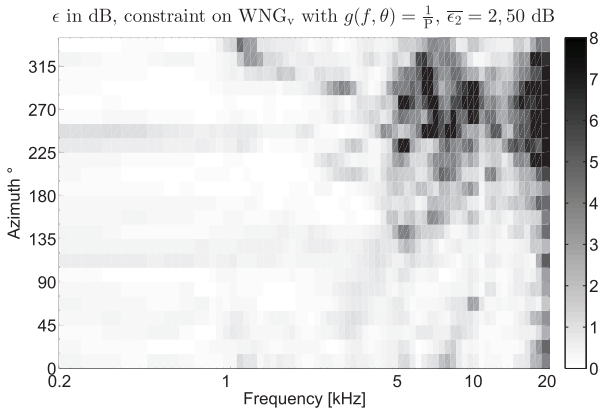


Fig. 10. Synthesis error ϵ as a function of frequency and direction for the synthesis of a left-ear HRTF associated with a constraint on $\mathrm{WNG_v}$ ($\beta_\mathrm{v} = -5$ dB).

This is also apparent in Fig. 9 depicting the synthesis error ϵ associated with a constraint on $\mathrm{WNG_m}$ as a function of frequency and direction. In general, the synthesis associated with a constraint on $\mathrm{WNG_m}$ yields a rather uniform and small error for ipsilateral directions ($0° \leq \theta \leq 180°$). However, this comes at the cost of a slightly larger error for contralateral directions ($180° \leq \theta \leq 360°$) compared to the synthesis associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$, especially at frequencies 5 kHz $\leq f \leq$ 10 kHz and $f \geq$ 15 kHz. The synthesized impulse responses associated with constraints on $\mathrm{WNG_m}$ and $\mathrm{WNG}(\mathbf{w}(f), 0°)$ are also depicted in Fig. 7. Both synthesized impulse responses resemble the desired head-related impulse responses quite well, with the largest deviations for the synthesized impulse responses associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$ and direction $\theta = 0°$.

Overall, the filter optimization associated with a constraint on $\mathrm{WNG_m}$ results in a much better synthesis compared to the filter optimization associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), \theta_d)$, in particular for the direction θ_d.

E. Constraint on the White Noise Gain for Multiple Frequencies

Fig. 10 depicts the synthesis error ϵ associated with a constraint on $\mathrm{WNG_v}(\mathbf{w_v}(\mathbf{\Omega}^t))$, incorporating the grouping of frequencies within ERBs as discussed in Section IV-C. As can be expected, the error is smoother and decreases slightly in comparison to the error associated with a constraint on $\mathrm{WNG_m}$ (cf. Fig. 9). This effect can be mainly observed for contralateral directions and for higher frequencies, which are associated with broader ERBs.

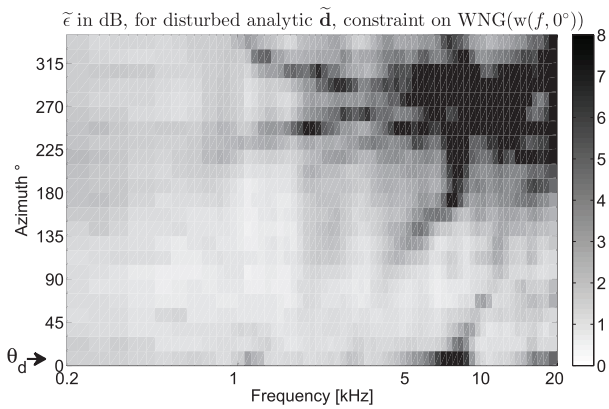


Fig. 11. Average synthesis error $\widetilde{\epsilon}$ for 100 randomly disturbed analytic steering vectors $\widetilde{\mathbf{d}}$ as a function of frequency and direction associated with a constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$.

F. Robustness of the Different Regularization Approaches

In Sections V-C to V-E, the synthesis error was investigated in case of no deviations from the measured steering vectors. In order to investigate the robustness against deviations of the steering vectors for the different regularization approaches, in this section we will analyze the synthesis error for disturbed steering vectors. This analysis will be performed for the measured as well as for analytic steering vectors to cover a broader variation of possible steering vectors. The analytic steering vectors were simulated assuming far-field conditions and assuming the same microphone array topology as the measured steering vectors. For both the measured and the analytic steering vectors, independent normally distributed vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ with zero-mean and variance $\sigma^2 = 1 - 10^{\frac{-0.05}{10}}$ (corresponding to 0.05 dB) were added to the real and the imaginary part of the steering vector $\mathbf{d}$ for each θ and f independently[5] (cf. [36]), resulting in the disturbed steering vector $\widetilde{\mathbf{d}}(f, \theta)$, i.e.

$$\widetilde{\mathbf{d}}(f,\theta) = (\Re\{\mathbf{d}(f,\theta)\} + \mathbf{r}_1(f,\theta)) + i \cdot (\Im\{\mathbf{d}(f,\theta)\} + \mathbf{r}_2(f,\theta)), \tag{29}$$

with $\Re\{\mathbf{d}\}$ and $\Im\{\mathbf{d}\}$ denoting the real and the imaginary part of the (undisturbed) steering vector $\mathbf{d}$, respectively.

A Monte-Carlo simulation with 100 realizations of the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ was performed. Fig. 11 depicts the average synthesis error $\widetilde{\epsilon}$ (averaged over 100 realizations of disturbed analytic steering vectors $\widetilde{\mathbf{d}}$) associated with a constraint on the white noise gain for a single direction $\mathrm{WNG}(\mathbf{w}(f), 0°)$. Although for $\theta = 0°$ this average synthesis error is quite similar to the synthesis error ϵ for the undisturbed measured steering vectors $\mathbf{d}$ in Fig. 8, the average synthesis error $\widetilde{\epsilon}$ in Fig. 11 is significantly larger, primarily for contralateral directions.

[5]Although these disturbances do not perfectly cover the real-world deviations of the steering vectors, they represent a large variety of real-world disturbances (e.g., sensor noise, positioning errors) and can hence be considered quite realistic.

TABLE I
MEAN SYNTHESIS ERRORS $\overline{\epsilon_1}$ AND $\overline{\epsilon_2}$ FOR THE PRESENTED REGULARIZATION APPROACHES WITHOUT AND WITH ADDITIONAL RANDOM DEVIATIONS OF THE ANALYTIC AND MEASURED STEERING VECTORS

$\overline{\epsilon_1}$ in dB	analytic		measured		$\overline{\epsilon_2}$ in dB	analytic		measured	
0 Hz $< f \leq$ 20 kHz	$\mathbf{d}$	$\widetilde{\mathbf{d}}$	$\mathbf{d}$	$\widetilde{\mathbf{d}}$	5 kHz $\leq f \leq$ 18 kHz	$\mathbf{d}$	$\widetilde{\mathbf{d}}$	$\mathbf{d}$	$\widetilde{\mathbf{d}}$
non-regularized	0.12	42.84	0	18.67	non-regularized	0	9.98	0	7.45
joint optimization with K=4	/	/	**1.06**	13.86	joint optimization with K=4	/	/	2.40	5.39
constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$	**1.52**	2.97	1.07	2.76	constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$	**2.69**	4.94	**2.18**	4.68
constraint on $\mathrm{WNG_m}$	1.70	2.76	1.22	2.64	constraint on $\mathrm{WNG_m}$	3.16	4.38	2.55	4.26
constraint on $\mathrm{WNG_v}$ with $g = \frac{1}{\mathrm{P}}$	1.65	**2.75**	1.20	**2.63**	constraint on $\mathrm{WNG_v}$ with $g = \frac{1}{\mathrm{P}}$	3.03	**4.35**	2.50	**4.21**

To obtain a single synthesis error, the mean synthesis error $\overline{\epsilon}$ was calculated over all realizations, all directions and all frequencies, i.e.

$$\overline{\epsilon} = \frac{1}{L_\epsilon}\frac{1}{P} \cdot \sum_{l=1}^{L_\epsilon}\sum_{i=1}^{P} \widetilde{\epsilon}(f_l, \theta_i), \tag{30}$$

with P the number of directions and L_ϵ the number of frequencies. Note that in case of undisturbed steering vectors, the average synthesis error $\widetilde{\epsilon}$ in (30) is equal to the synthesis error ϵ. For the mean synthesis error we consider two different frequency ranges. The mean synthesis error $\overline{\epsilon_1}$ was calculated for the complete frequency range, but excluding the DC, i.e. for frequencies $f > 0$ Hz with f starting at the first frequency bin, i.e. fs/NFFT $\approx$ 86 Hz. In contrast to the mean synthesis error $\overline{\epsilon_1}$ the mean synthesis error $\overline{\epsilon_2}$ is limited to frequencies 5 kHz $\leq f \leq$ 18 kHz. This frequency band was chosen to cover the most challenging frequencies ($f > 5$ kHz) when synthesizing HRTFs. The upper limit of 18 kHz was chosen with the aim to include the highest frequencies which are still audible at reasonable threshold levels of young normal hearing adults. Therefore, we deemed the synthesis for $f > 18$ kHz to be less relevant from a perceptual point of view while at the same time these very high frequencies exhibit very large errors (cf. Figs. 8–11) which would dominate the mean synthesis error. Hence, $\overline{\epsilon_1}$ covers the entire frequency band (excluding DC) and $\overline{\epsilon_2}$ covers the most challenging yet perceptually relevant frequency range.

First, from the results in Table I it can be observed that both mean synthesis errors ($\overline{\epsilon_1}$ and $\overline{\epsilon_2}$) increase for disturbed steering vectors $\widetilde{\mathbf{d}}$ compared to undisturbed steering vectors $\mathbf{d}$, both for analytic as well as for measured steering vectors and for all (regularization) approaches. As expected, for the non-regularized filter optimization the mean synthesis errors increase drastically when the steering vectors are disturbed.

For the joint optimization (using $\mathrm{K} = 4$ sets of measured steering vectors) the mean synthesis errors for the disturbed steering vectors are only slightly smaller than for the non-regularized optimization. Hence, the joint optimization with $\mathrm{K} = 4$ only slightly enhances the robustness of the synthesis, which may presumably be enhanced by incorporating more sets of measured steering vectors. The synthesis errors associated with the joint optimization indicate that this regularization approach is not well suited for variations of the steering vectors that were not considered in the optimization. Since in real life applications it is hardly possible to measure and consider all possible variations of the steering vectors, this regularization approach is considered to be disadvantageous in comparison to the other proposed regularization approaches.

Out of all considered regularization approaches, the optimization associated with a constraint on the white noise gain for a single direction $\mathrm{WNG}(\mathbf{w}(f), 0°)$ yields the smallest mean synthesis errors when assuming undisturbed analytic steering vectors and the smallest $\overline{\epsilon_2}$ when assuming undisturbed measured steering vectors. This may be explained by the direction-dependent regularization which primarily influences only the direction $\theta_d = 0°$. When considering disturbed steering vectors, a constraint on the white noise gain for a single direction yields smaller mean synthesis errors than the joint optimization with $\mathrm{K} = 4$ (but larger mean synthesis errors compared to constraints on $\mathrm{WNG_m}$ and $\mathrm{WNG_v}$). However, it should be kept in mind that a white noise gain for a single direction yields a direction-dependent impact on the synthesis (cf. Section V-C) and is hence undesirable for a multi-directional synthesis.

For the constraints on the mean white noise gains $\mathrm{WNG_m}$ and $\mathrm{WNG_v}$, the mean synthesis errors $\overline{\epsilon_1}$ and $\overline{\epsilon_2}$ are clearly smaller for the disturbed steering vectors compared to the constraint on $\mathrm{WNG}(\mathbf{w}(f), 0°)$ and the joint optimization with $\mathrm{K} = 4$. This holds for the analytic as well as for the measured steering vectors, showing that these proposed regularization approaches improve consistently robustness. The highest robustness is obtained by the constraint on the mean white noise gain $\mathrm{WNG_v}$ taking into account multiple frequencies.

In conclusion, the proposed regularization approaches constraining the mean white noise gains $\mathrm{WNG_m}$ and $\mathrm{WNG_v}$ seem to be appropriate for synthesizing head-related transfer functions for a virtual artificial head. The multi-directional synthesis accuracy and robustness is clearly enhanced when considering all directions for the white noise gain ($\mathrm{WNG_m}$) and when incorporating a psychoacoustically-motivated frequency grouping ($\mathrm{WNG_v}$).

VI. CONCLUSION

In this paper, several approaches to increase the robustness of a filter-and-sum beamformer for synthesizing multi-directional spatial directivity patterns were presented, with a focus on synthesizing HRTFs using a virtual artificial head.

Firstly, an optimization procedure incorporating multiple measured steering vectors was shown to enhance the robustness against random deviations of the steering vectors compared to a non-regularized optimization. However, a joint optimization for only 4 different measured sets of steering vectors resulted in a relatively low robustness at low frequencies. In general this regularization approach seems primarily appropriate in theory

and impracticable for real life applications, since typically not enough variations of the steering vectors can be measured.

Secondly, a mean weighting of the WNG over all directions was presented and shown to outperform the WNG for a single look direction when synthesizing multi-directional spatial directivity patterns.

Thirdly, a design procedure incorporating multiple frequency bins in the optimization and regularization was presented. It was shown that the approach incorporating frequencies within ERBs resulted in the best robustness of the synthesis, for measured as well as for analytic steering vectors. This shows the suitability of the presented regularization approaches for synthesizing HRTFs using microphone arrays.

REFERENCES

[1] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 94, no. 1, pp. 111–123, Jul. 1993.

[2] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," *J. Acoust. Soc. Amer.*, vol. 99, no. 6, pp. 3678–3688, Jun. 1996.

[3] P. A. Hill, P. A. Nelson, O. Kirkeby, and H. Hamada, "Resolution of front-back confusion in virtual acoustic imaging systems," *J. Acoust. Soc. Amer.*, vol. 108, no. 6, pp. 2901–2910, Dec. 2000.

[4] B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[5] S. Doclo, W. Kellermann, S. Makino, and S. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[6] H. Cox, R. Zeskind, and T. Kooij, "Practical supergain," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 34, no. ASSP-34, pp. 393–398, Jun. 1986.

[7] J. M. Kates, "Superdirective arrays for hearing aids," *J. Acoust. Soc. Amer.*, vol. 94, no. 4, pp. 1930–1933, Oct. 1993.

[8] J. Bitzer and K. U. Simmer, "chapter 2 in "Microphone arrays: Signal processing techniques and applications"," in *Superdirective Microphone Arrays*, M. S. Brandstein and D. B. Ward, Eds. Berlin, Germany: Springer-Verlag, May 2001, pp. 19–38.

[9] S. Doclo and M. Moonen, "Superdirective beamforming robust against microphone mismatch," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 2, pp. 617–631, Feb. 2007.

[10] E. Mabande, A. Schad, and W. Kellermann, "Design of robust superdirective beamformers as a convex optimization problem," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2009, pp. 77–80.

[11] S. Nordebo, I. Claesson, and S. Nordholm, "Weighted Chebyshev approximation for the design of broadband beamformers using quadratic programming," *IEEE Signal Process. Lett.*, vol. 1, no. 7, pp. 103–105, Jul. 1994.

[12] M. Kajala and M. Hämäläinen, "Broadband beamforming optimization for speech enhancement in noisy environments," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 1999, pp. 19–22.

[13] S. Doclo and M. Moonen, "Design of far-field and near-field broadband beamformers using eigenfilters," *Signal Process.*, vol. 83, no. 12, pp. 2641–2673, Dec. 2003.

[14] J. Chen, B. D. V. Veen, and K. E. Hecox, "External ear transfer function modeling: A beamforming approach," *J. Acoust. Soc. Amer.*, vol. 92, no. 4, pp. 1933–1944, Oct. 1992.

[15] Y. Kahana, P. A. Nelson, O. Kirkeby, and H. Hamada, "A multiple microphone recording technique for the generation of virtual acoustic images," *J. Acoust. Soc. Amer.*, vol. 105, no. 3, pp. 1503–1516, Mar. 1999.

[16] N. Tohtuyeva and V. Mellert, "Approximation of dummy-head recording technique by a multimicrophone arrangement," *J. Acoust. Soc. Amer.*, vol. 105, no. 2, p. 1101, 1999.

[17] S. Sakamoto, R. Kadoi, S. Hongo, and Y. Suzuki, "SENZI and ASURA: New high-precision sound-space sensing systems based on symmetrically arranged numerous microphones," in *Proc. 2nd Int. Symp. Universal Commun. (ISUC'08)*, 2008, pp. 429–434.

[18] S. Sakamoto, S. Hongo, and Y. Suzuki, "3D sound-space sensing system based on numerous symmetrically arranged microphones," *IEICE Trans. Fundam. Electron., Commun., Comput. Sci.*, vol. E97-A(9), pp. 1893–1901, 2014.

[19] E. Rasumow, M. Blau, M. Hansen, S. Doclo, S. van de Par, V. Mellert, and D. Püschel, "Robustness of virtual artificial head topologies with respect to microphone positioning errors," in *Proc. Forum Acusticum*, Aalborg, Denmark, Jul. 2011, pp. 2251–2256.

[20] J. Atkins, "Spatial acoustic signal processing for immersive communication," Ph.D. dissertation, Johns Hopkins Univ., Baltimore, MD, USA, Aug. 2011.

[21] J. Atkins, "Robust beamforming and steering of arbitrary beam patterns using spherical arrays," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 2011, pp. 237–240.

[22] E. Rasumow, M. Blau, S. Doclo, M. Hansen, S. Van de Par, D. Püschel, and V. Mellert, "Least squares versus non-linear cost functions for a virtual artificial head," in *Proc. Meetings Acoust.*, Jun. 2013, vol. 19, ASA.

[23] E. Rasumow, M. Blau, M. Hansen, S. Doclo, S. van de Par, V. Mellert, and D. Püschel, "The impact of the white noise gain (WNG) of a virtual artificial head on the appraisal of binaural sound reproduction," in *Proc. EAA Joint Symp. Auralization Ambisonics*, Berlin, Germany, Apr. 2014, pp. 174–180.

[24] S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2511–2526, Oct. 2003.

[25] D. Levin, E. A. P. Habets, and S. Gannot, "Robust beamforming using sensors with nonidentical directivity patterns," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 91–95.

[26] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, D.C.: Winston & Sons, 1977.

[27] J. L. Castellanos, S. Gomez, and V. Guerra, "The triangle method for finding the corner of the L-curve," *Appl. Numer. Math.*, vol. 43, no. 4, pp. 359–373, Dec. 2002.

[28] P. C. Hansen, "Regularization tools - A Matlab package for analysis and solution of discrete ill-posed problems," [Online]. Available: http://www.imm.dtu.dk/~pch date accessed 16/12/15

[29] H. Fletcher, "Auditory patterns," *Rev. Mod. Phys.*, vol. 12, pp. 47–65, 1940.

[30] B. C. J. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Amer.*, vol. 74, no. 3, pp. 750–753, Sep. 1983.

[31] E. Rasumow, M. Blau, M. Hansen, S. van de Par, S. Doclo, V. Mellert, and D. Püschel, "Smoothing individual head-related transfer functions in the frequency and spatial domains," *J. Acoust. Soc. Amer.*, vol. 135, no. 4, pp. 2012–2025, Apr. 2014.

[32] L. D. Mitchell, "Improved methods for the fast Fourier transform (FFT) calculation of the frequency response function," *J. Mech. Design*, vol. 104, pp. 277–279, Apr. 1982.

[33] D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," *J. Acoust. Soc. Amer.*, vol. 100, no. 1, pp. 408–427, 1996.

[34] M. Omura, M. Yada, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Compensating of room acoustic transfer functions affected by change of room temperature," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 1999, vol. 2, pp. 941–944.

[35] Y. Yasuno and J. Ohga, "Temperature characteristics of electret condenser microphones," in *Proc. Int. Symp. Electrets*, Sep. 2005, pp. 412–415.

[36] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*. New York, NY, USA: Wiley, 1986, pp. 74–109, Chapter 4.

[37] E. Rasumow, M. Blau, S. Doclo, S. van de Par, M. Hansen, D. Püschel, and V. Mellert, "Perceptual evaluation of individualized binaural reproduction using a virtual artificial head," *PlosOne*, Sep. 2015, submitted for publication.

[38] E. Rasumow, "Synthetic reproduction of head-related transfer functions by using microphone arrays," Ph.D. dissertation, Univ. of Oldenburg, Oldenburg, Germany, Mar. 2015.

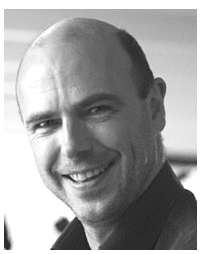
Eugen Rasumow received the Dipl.-Ing. (FH) degree in hearing technology and audiology from the Oldenburg University of Applied Science, Germany, in 2007 and the M.Sc. degree in hearing technology and audiology from the University of Oldenburg, Germany, in 2009. In 2015 he received the Ph.D. degree from the University of Oldenburg, Germany.

His research interests include psychoacoustics, virtual acoustics and binaural signal processing.

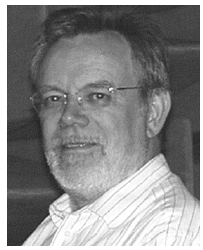
Martin Hansen studied physics in Göttingen and did his Ph.D. in Oldenburg about psychoacoustical modeling of low bit-rate coded speech transmission quality. After working in R&D in a hearing aid company in Copenhagen for several years he became Professor for medical acoustics at the Jade University for Applied Sciences in Oldenburg.

His research interests include psychoacoustics and technical audiology.

Steven van de Par studied physics at the Eindhoven University of Technology, Eindhoven, The Netherlands, and received the Ph.D. degree from the Eindhoven University of Technology, in 1998, on a topic related to binaural hearing. As a Postdoctoral Researcher at the Eindhoven University of Technology, he studied auditory-visual interaction and was a Guest Researcher at the University of Connecticut Health Center. In early 2000, he joined Philips Research, Eindhoven, to do applied research in digital signal processing and acoustics. His main fields of expertise are auditory and multisensory perception, low-bit-rate audio coding and music information retrieval. He has published various papers on binaural auditory perception, auditory-visual synchrony perception, audio coding, and music information retrieval (MIR)-related topics. Since April 2010, he has held a professor position in acoustics at the University of Oldenburg, Oldenburg, Germany.

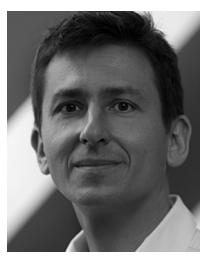


Dirk Püschel was born in 1960. He studied physics from 1980–1984 and received his Ph.D. degree in 1988. His research was in the field of psychoacoustic and roomacoustic at III. Institute of Physics in Goettingen until 1993. He was the founder of Akustikbüro Goettingen in 1992 and Akustik Technologie Goettingen in 1993. He is the Owner, Director of Research, and Managing Director of Soundtec GmbH.



Volker Mellert received his Ph.D. in physics in Göttingen 1972, and became a Professor for Applied Physics in Oldenburg in 1976. He was head of the Acoustics Group at the Institute for Physics of Oldenburg University, and retired in 2009. Acoustic wave propagation and human perception (psychoacoustics) are his main scientific interests. He was President of the German Association for Acoustics (DEGA), and of the European Association for Acoustics (EAA). He served as board member of the International Commission of Acoustics, and as Editor-in-Chief for ACUSTICA. He received the German Medal of Merit for his contributions in acoustics and the Helmholtz-Medal of the DEGA. He is Fellow of the Acoustical Society of America.

Simon Doclo (S'95–M'03–SM'13) received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in 1997 and 2003. From 2003 to 2007 he was a Postdoctoral Fellow with the Research Foundation Flanders at the Electrical Engineering Department (Katholieke Universiteit Leuven) and the Adaptive Systems Laboratory (McMaster University, Canada). From 2007 to 2009 he was a Principal Scientist with NXP Semiconductors at the Sound and Acoustics Group in Leuven, Belgium. Since 2009 he has been a Full Professor at the University of Oldenburg, Germany, and Scientific Advisor for the project group Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology. His research activities center around signal processing for acoustical applications, more specifically microphone array processing, active noise control, acoustic sensor networks and hearing aid processing. Prof. Doclo received the Master Thesis Award of the Royal Flemish Society of Engineers in 1997 (with Erik De Clippel), the Best Student Paper Award at the International Workshop on Acoustic Echo and Noise Control in 2001, the EURASIP Signal Processing Best Paper Award in 2003 (with Marc Moonen) and the IEEE Signal Processing Society 2008 Best Paper Award (with Jingdong Chen, Jacob Benesty, Arden Huang). He was member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing (2008-2013) and Technical Program Chair for the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2013. Prof. Doclo has served as guest editor for several special issues *(IEEE Signal Processing Magazine*, *Elsevier Signal Processing*) and is associate editor for IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING and *EURASIP Journal on Advances in Signal Processing*.



Matthias Blau studied electrical engineering at TU Dresden (Germany) and Purdue University (USA). In 1999 he received a Ph.D. in electrical engineering from TU Dresden. Since 2003 he has been a Professor of Electroacoustics at Jade Hochschule in Oldenburg, Germany.

His current research interests include the acoustics of outer and middle ears, virtual acoustics, implantable transducers, room acoustics, and acoustical measurements.

Dr. Blau is member of the German Acoustical Society, the Acoustical Society of America, and the Marie Curie Fellowship Association.